Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com

When robots tell you what to do: Sense of agency in human- and robot-guided actions

Zeynep Barlas[a,b]

[a] Social Cognitive Systems – Cluster of Excellence Center in Cognitive Interactive Technology (CITEC), Bielefeld University, Bielefeld, Germany
[b] Institute for Psychological Science, School of Applied Social Sciences, De Montfort University, Leicester, United Kingdom

ARTICLE INFO

Keywords:
Sense of agency
Intentional binding
Action selection
Humanoid robots
Robot autonomy
Artificial agents

ABSTRACT

The present study investigated the sense of agency (SoA) when actions were determined by another human vs. a humanoid robot as compared to when freely selected. Additionally, perceived robot-autonomy was manipulated via autonomous vs. non-autonomous descriptions of the robot. SoA was assessed by judgment of control ratings and intentional binding (i.e., perceived temporal attraction between voluntary actions and their outcomes). Participants performed free and instructed key presses that produced an auditory tone (Experiment-1) and visual stimuli conveying neutral, positive, or negative valence (Experiment-2). Binding and control ratings were greater in free compared to instructed actions, and comparable between human- and robot-instructed actions. Control ratings were higher for positive compared to neutral and negative outcomes, and positively correlated with ratings of how human-like the robot appeared. These results highlight the role of endogenous processing of action selection and provide preliminary insight into the SoA when actions are guided by artificial agents.

1. Introduction

In the last a few decades, artificial intelligence technology has dramatically been advanced with the development of more sophisticated artificial systems such as virtual agents and humanoid robots that accompany or substitute humans in many industrial and daily tasks. Indeed, today various robots appear as companions to humans in a plethora of roles in education, cleaning, physical activities, cognitive tasks, and health care (Broadbent, 2017). As this technological development has led to more frequent interactions between humans and artificial agents, psychological research has consequently confronted a novel area wherein human experience in these interactions is investigated. The present paper is concerned with how one's experience of control over their actions could be altered in interactions with artificial systems, specifically when the interaction involves individuals receiving action instructions from a humanoid robot and accordingly performing these instructed actions.

From daily routine actions to more skilful and complex actions, we experience some degree of the sense that we are in control of our actions and the consequences of these actions (Gallagher, 2000; Haggard & Chambon, 2012; Haggard & Tsakiris, 2009). This experience, namely the sense of agency (SoA), bears strong connections to one's feeling of responsibility and has important implications for societal and legal settings (Caspar, Christensen, Cleeremans, & Haggard, 2016; Caspar, Cleeremans, & Haggard, 2018; Haggard, 2017; Moretto, Walsh, & Haggard, 2011). As such, the critical question regarding how SoA is emerged and altered under certain circumstances has recently found grounds in psychological and neuroscientific research. It has been proposed that SoA is emerged as a result of a weighted combination of several factors, depending on their reliability and involvement in prospective (i.e.,

E-mail address: zbarlas@techfak.uni-bielefeld.de, zeynep.barlas@dmu.ac.uk (Z. Barlas)

https://doi.org/10.1016/j.concog.2019.102819
Received 27 April 2019; Received in revised form 17 July 2019; Accepted 8 September 2019
Available online xxx
1053-8100/

before the movement) and retrospective (i.e., after the movement) processes (Moore & Fletcher, 2012). For instance, several studies have reported that individuals experience weaker control when there is a discrepancy between sensorimotor predictions and actual outcomes (e.g., Barlas & Kopp, 2018; Ebert & Wegner, 2010; Sato & Yasuda, 2005). The role of sensorimotor predictions produced by internal forward models (Wolpert, 1997; Wolpert, Ghahramani, & Jordan, 1995) and their congruency with actual sensory consequences of actions on SoA has strongly been emphasized in the so-called Comparator Model of motor learning and control mechanisms (Blakemore, Wolpert, & Frith, 2002; Frith, 2005; Frith, Blakemore, & Wolpert, 2000). SoA in individual actions has also been shown to be vulnerable to temporal contiguity between actions and their outcomes (Moore, Wegner, & Haggard, 2009; Wegner & Wheatley, 1999). Unpredicted action-outcome delays, for instance, can retrospectively inform the SoA, yielding stronger experience of control after shorter compared to longer delays between actions and their corresponding effects (Barlas, Hockley, & Obhi, 2018; Chambon & Haggard, 2012; Ebert & Wegner, 2010; Wen, Yamashita, & Asama, 2015). Interestingly, our experience of agency can also modulate our time perception of the events surrounding our actions. More specifically, in voluntary compared to involuntary actions, perceived times of actions are shifted towards the perceived times of outcomes, and vice versa, yielding a perceived temporal binding of actions and their outcomes (Haggard, Clark, & Kalogeras, 2002). The temporal attraction between self-generated actions and their outcomes is hailed as the intentional binding phenomenon (Haggard et al., 2002). As an alternative or in addition to employing explicit measures that acquire self-reports of how much control is experienced under specific experimental conditions (Barlas et al., 2018; Sidarus, Vuorre, & Haggard, 2017; Wenke, Fleming, & Haggard, 2010), assessing the degree of intentional binding has been extensively used as an implicit measure of the SoA (see Moore & Obhi, 2012 for a review).

As mentioned earlier, the engagement of artificial systems in human lives has dramatically increased. In this respect, a recently emerging question is how SoA could be altered during interactions and joint actions with artificial systems compared to during individual actions or joint actions with others (for relevant reviews see Limerick, Coyle, & Moore, 2014; Sahaï, Pacherie, Grynszpan, & Berberian, 2017). Arguably, addressing this question could provide further insight into understanding the underpinnings of SoA and potentially modulate the design of artificial systems to increase their efficacy and acceptability. Previous research approached this issue from different perspectives. Berberian and colleagues, for instance, examined intentional binding and subjective control ratings during an aircraft supervision task using a flight simulator system (Berberian, Sarrazin, Le Blaye, & Haggard, 2012). In their study, participants were required to implement the appropriate command to resolve the conflict that could emerge due to the presence of another plane. After participants implemented the command by pressing a key, they were provided with a visual and auditory feedback and required to estimate the delay between their key press and the feedback. Importantly, participants involved in the task at varying levels between having full control over this decision and merely observing the simulation system performing the task with full autonomy. For each level of system-autonomy, they also indicated how strongly they felt causal control in the task. The results showed that both intentional binding and control ratings were reduced as the system was bestowed with more autonomy to complete the task. Conversely, in human-human joint actions, it might possible to experience a form of we-agency when individuals perform cooperative joint actions (Obhi & Hall, 2011a, 2011b; Strother, House, & Obhi, 2010). Obhi and Hall (2011b), for instance, demonstrated that comparable intentional binding was observed for self-caused outcomes and those that were believed to be caused by a human partner's actions. However, when participants believed that the interaction partner was a computer, intentional binding for both their own and the computer's actions was diminished. That is, an implicit level of SoA indexed by intentional binding could only be exhibited when the interaction partner was another human in contrast to an artificial agent such as a computer. Conversely, another study showed that repeated experience with observing another person and a human-like robotic hand performing the same action yielded similar progression of temporal binding between one's actions and the corresponding outcomes (Khalighinejad, Bahrami, Caspar, & Haggard, 2016).

It can be conceived based on these results that SoA is distinctively experienced in joint actions depending on the identity of the action partner. One view in this respect suggests that reduced SoA in joint actions with artificial systems as compared to with other humans may be due to the failure to co-represent the intentions and motor plans of these artificial systems (Obhi & Hall, 2011b; Wohlschläger, Engbert, & Haggard, 2003). In line with this view, Sahaï et al. (2017) contended that predictive mechanisms could function distinctively between human and machine action partners, hence altered SoA depending on the identity of the action partner. Nevertheless, there is a considerable variety among artificial systems from very simplistic ones (e.g., computers) to those resembling humans in many physical aspects (e.g., humanoid robots). Correspondingly, another contention is that human-like features as well as high level beliefs about intentionality and agentic capacity of artificial agents could also take a critical role in determining how humans co-represent the actions of these agents (Stenzel et al., 2012, 2014). In Stenzel et al. (2012), for instance, believing that a humanoid robot was an autonomous and biologically inspired agent yielded the Social Simon effect[1] (Sebanz, Knoblich, & Prinz, 2003), the presence of which implies that the actions of the robot partner were co-represented by the participants. Interestingly, however, this effect was diminished when the robot was introduced as a machine-like and deterministic agent to the participants. In a similar vein, a recent study using a computer and another person as the action partners in a Simon task

[1] Social Simon task is a well-established paradigm to assess co-representation of others' actions. In a standard Simon task (Simon, 1969; Simon & Rudell, 1967), participants perform one of two spatially defined actions (e.g., right and left key presses) in response to the corresponding stimulus (e.g., a square and a diamond) that is presented on a spatially congruent or incongruent position in relation to the response (e.g., right or left side of the screen). The Simon effect refers to that responses are faster when the stimulus position spatially matches the corresponding response action. While this effect is not observable if the participant has to respond to only one of the two stimuli (Hommel, 1996), it re-emerges when the participant is paired with another person who is to respond to the alternative stimulus. The Social Simon effect thus suggests that in joint actions, individuals automatically co-represent each other's actions (Sebanz et al., 2003).

found that both intentional binding and Social Simon effect were diminished when participants were paired with a computer in comparison to when cooperating with another person (Sahaï, Desantis, Grynszpan, Pacherie, & Berberian, 2019).

Taken together, previous research examining SoA in interactions with different artificial systems suggests that performing actions in collaboration with these agents might impair one's SoA as compared to individual and human-human joint actions (Berberian et al., 2012; Obhi & Hall, 2011b; Sahaï et al., 2019), while human-like features of the artificial partner could also play a critical role (Khalighinejad, Bahrami, et al., 2016). It should be noted that majority of previous studies involved joint action scenarios in which participants and artificial agents either jointly or alternately performed the experimental tasks. In many applications of artificial agents, however, the major role of these agents is to guide human activities. Perhaps the most commonly recognized application as such is the navigation systems, which are frequently used to get directions towards which to steer to reach a specific destination. Another example is the employment of humanoid robots at shopping malls in Japan. These robots not only provide information about the mall but also give directions and recommendations for shopping (Satake, Hayashi, Nakatani, & Kanda, 2015). Additionally, iRobi and Cafero (Yujin Robot) are two assistive robots that remind patients to take medications and provide cognitive stimulation (Broadbent, 2017). Little is known, however, how one's SoA would be altered in such interactions with artificial agents that determine which action to take compared to when one determines their actions themselves or receives guidance by another person. A few recent studies investigated the effect of the source of actions selection (i.e., free vs. externally determined) on SoA and showed that one's experience of agency is weakened when one performs an externally selected (by a computer or by another human) compared to when one freely determines what to do (e.g., Barlas et al., 2018; Caspar et al., 2016, 2018). Furthermore, the effect of free selection on the SoA was found independent of the identity of action-outcomes, and enhanced with an increased number of action choices (Barlas & Kopp, 2018; Barlas & Obhi, 2013; Barlas et al., 2018; Barlas, Hockley, & Obhi, 2017).

The primary goal of the current study was to examine how receiving action instructions from a humanoid robot as compared to a human would influence intentional binding and explicit judgment of control over one's actions as also compared to when actions are freely selected. Second, the current study aimed to investigate whether believing that a humanoid robot is autonomous and capable of making purposeful action decisions or not could differentially affect one's SoA when performing robot-instructed actions. Accordingly, in two experiments, intentional binding and judgment of control were assessed while participants performed free choice actions and actions instructed by another person or a humanoid robot. Additionally, the belief about robot-autonomy was manipulated by introducing the robot to the participants as either autonomous or non-autonomous. Autonomy of the robot in this context particularly refers to the ability to make its own decisions and give purposeful action instructions. In Experiment 1, all actions (right and left key presses) produced the same auditory outcome while in Experiment 2, these actions produced visual outcomes (face stimuli) with different expressions that conveyed neutral, positive, or negative valence. At the end of both experiments, participants completed a questionnaire with several items (anthropomorphism, likeability, intelligence, perceived intentionality, and decision-making ability), that assessed how human-like and autonomous the robot was perceived by the participants.

For both experiments, it was predicted that participants in the autonomous group would report higher ratings of intentionality and decision-making ability compared to the non-autonomous group (Stenzel et al., 2012). Based on the previous results indicating enhanced binding and subjective control in freely selected compared to externally determined actions (Barlas & Obhi, 2013; Barlas et al., 2017, 2018; Caspar et al., 2016, 2018), stronger binding and experienced control were predicted when participants performed free compared to both human- and robot-instructed actions. It was also conjectured that if perceived autonomy of the robot determined how actions goals are represented and incorporated into one's motor control mechanisms, then human- and robot-instructed conditions were expected to yield comparable SoA when the robot was believed to be an autonomous compared to a non-autonomous agent.

Predictions towards the effect of outcome-valence on explicit control ratings (Experiment 2) were based on the notion of self-serving bias (Duval & Silvia, 2002; Miller & Ross, 1975; Taylor & Brown, 1994), which refers to the tendency to attribute oneself as the cause of positive as opposed to negative or undesirable events. Accordingly, control ratings were expected to be higher for positive compared to negative and neutral outcomes. In addition, negative outcomes were predicted to yield weaker experience of control compared to neutral outcomes. Importantly, each outcome-valence in Experiment 2 was randomly produced by each action and thus, the valence of outcomes was only retrospectively evident to the participants. Based on a recent view suggesting that the effect of valence on binding requires a predictive model of action-outcomes (Yoshie & Haggard, 2017), a similar degree of binding was expected across neutral, positive, and negative conditions.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Sample size was determined *a priori* using G*Power 3.1. (Faul, Erdfelder, Buchner, & Lang, 2009), which suggested 30 participants per group for a within-subjects effect of action-choice to achieve a power of 0.95 with a large effect size ($d = 0.80$, alpha = 0.05, Cohen, 1988). In total thus, 60 participants (30 per group) were recruited from Bielefeld University (33 females, 6 left-handed, $M_{age} = 24.83$ years, $SD = 4.67$ years). Participants were randomly assigned to one of the autonomous (n = 30, 16 females, 4 left-handed, $M_{age} = 24.93$ years, $SD = 4.55$ years) and non-autonomous groups (n = 30, 17 females, 2 left-handed, $M_{age} = 24.73$ years, $SD = 4.86$ years). All participants had normal or corrected-to-normal vision and had no hearing problems. Par-

ticipants gave their written consent prior to the study and received monetary compensation (10 Euros) in exchange for their participation. The study was conducted in accordance with the ethical guidelines of the Declaration of Helsinki and was approved by the Research Ethics Board of Bielefeld University.

2.1.2. Materials and experimental setup

The experiment was developed using PsychoPy v3.0.4 (Peirce, 2007, 2008). Software modules to control the robot behavior was developed using Python NAOqi (2.4.1). The humanoid robot used in the experiment was a NAO robot (Softbank-Aldebaran Robotics, https://www.softbankrobotics.com/emea/en/nao). All software was run on a Dell computer (3.07 GHz) and participants were seated approximately 75 cm away from a 20-in. monitor (resolution: 1600 × 1200). For the experimental conditions in which the robot was not involved, a separator box was placed on the desk between the participant's sitting area and the robot (see Fig. 1). The rationale for placing this separator was twofold. First, it helped ensure that the robot did not distract the participants in the passive, free, and human-instructed conditions. Second, in the human-instructed condition, the robot would blink right/left eye to provide the experimenter with the right/left instruction cue and the separator was helpful to keep this unbeknown to the participants. Visual stimuli were presented on a white-background screen and comprised a star sign ("*", 66 pt), a control rating scale, and an interval estimation scale. A standard keyboard was used for the key press responses. On this keyboard, "c" and "m" keys were labelled with "L" and "R" letter prints, that were associated with left and right key press actions, respectively. An optical wheel mouse was used to indicate responses on the interval estimation and control rating scales (Barlas & Kopp, 2018). The interval estimation scale was ranged from 1 to 1000 ms and marked at 50 ms intervals while the control rating scale was a 6-point Likert scale marked at 0.5-point intervals (1: very weak; 6: very strong). Both scales occupied 1000 pixels on the screen and the horizontal axis position of the mouse cursor after the response (at pixel-precision) was used for the conversion to the corresponding scale value (i.e., between 1

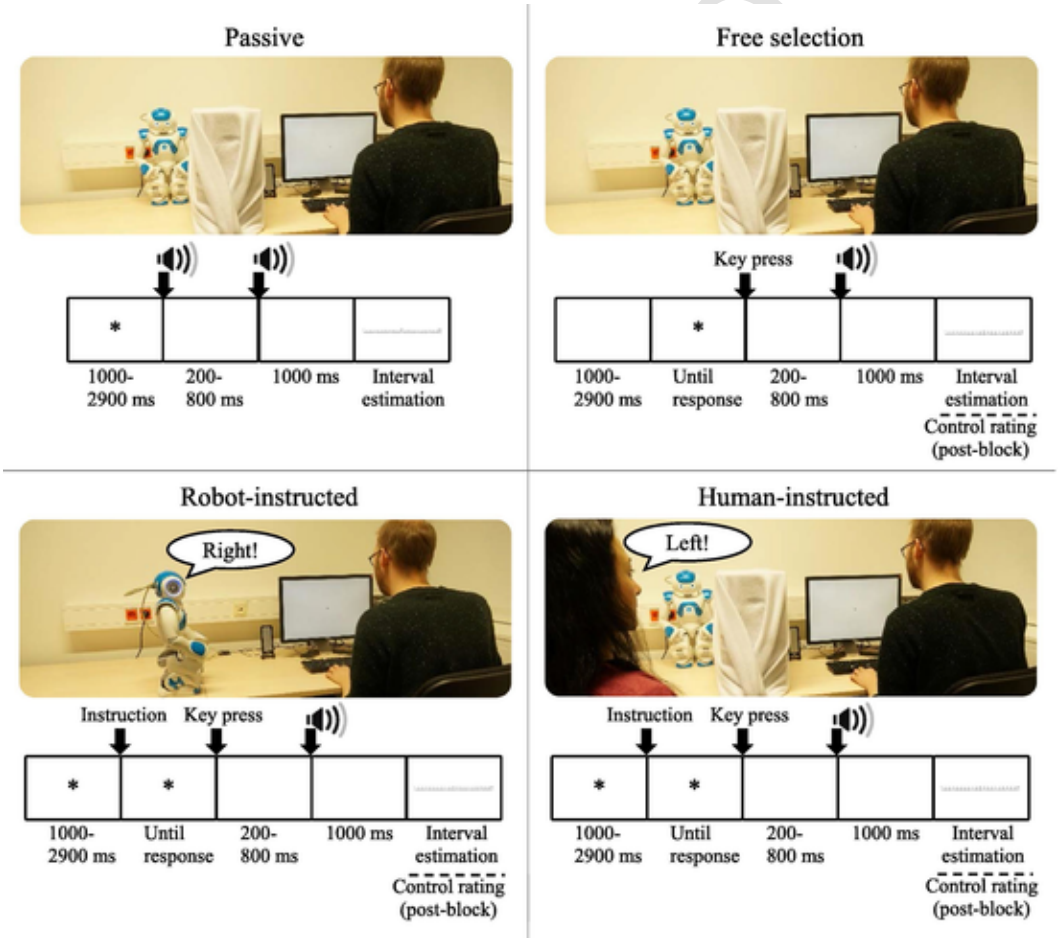

Fig. 1. Illustration of each condition and corresponding trial procedure. In the passive condition, participants estimated the delay between two passively heard sounds (a key press click sound and a beep sound). In the free condition, they freely chose between the right and the left key while in the human- and robot-instructed conditions, they pressed the instructed key at a time of their choice. At the end of each trial, they were asked to estimate the delay between their key press and the beep sound. At the end of each block except the passive condition, they also indicated how much control they experienced over the beep sound on a 6-point scale (1: very weak; 6: very strong). Additionally, participants completed a post-experiment questionnaire that assessed the robot in terms of anthropomorphology, likeability, perceived intelligence, and whether it appeared intentional and capable of making its own decisions.

and 1000 ms for the interval estimation scale and between 1 and 6 for the control rating scale). Participants used their left hand to perform the key press actions, and they used their right hand to control the mouse when indicating interval estimations and control ratings.

Auditory stimuli were generated using Audacity 2.2.1 (https://www.audacityteam.org/) and consisted of a beep sound (50 ms, bit rate: 705 kbps) and the sound of pressing a key (50 ms, bit rate: 1411 kbps). The key press sound was recorded using the same keyboard used in the experiment and was used only in the passive condition (see below). All auditory stimuli were delivered at 60 dB through the speakers located on each side of the computer.

2.1.3. Procedure

The duration of the experiment was approximately 60 min. Participants first read and signed the informed consent form and completed the demographic questionnaire that obtained information regarding age, gender, and handedness. The experimenter then briefly informed about the goal of the experiment and introduced the apparatus. Specific instructions regarding the experimental task were given at the beginning of each condition. Experimental design comprised action-choice (passive, free, human-instructed, and robot-instructed) and action-outcome interval (200 ms, 400 ms, 600 ms, 800 ms) as the within-subjects variables and robot-autonomy (autonomous, non-autonomous) as the between-subjects variable. Action-choice was presented in four blocks while action-outcome interval was mixed in each block. Robot-autonomy was determined by how the robot, named Zora, was introduced at the beginning of the robot-instructed condition. Accordingly, the autonomous group was told that Zora could make its own decisions by modeling how humans determine their actions and thus, Zora would actively decide in each trial which key participants should press. The non-autonomous group was informed that Zora's key press instructions were pre-programmed, and Zora would simply tell them which key to press. The order of action-choice blocks was counterbalanced across participants using a 4 × 4 Latin Square procedure (participants after a sample size of 24 were randomly assigned with one order). In total, participants completed 320 experimental trials (80 per block) in the interval estimation task. Additionally, 5 practice trials were presented at the beginning of each block. The experiment paused in the middle of each block to give a short break and participants pressed one of the two (left or right) keys to resume the experiment when they were ready.

All trials started with a 250 ms presentation of "Next". In the passive condition, this was followed by the display of a star sign ("*") which disappeared after a random delay (jittered between 1000 and 2900 ms) simultaneously with a key press sound. The purpose of presenting a key press sound was to render the both visual and auditory stimuli in the passive condition comparable to the conditions in which participants performed a key press. Disappearance of the star sign was followed by a blank screen presented for a random delay (200 ms, 400 ms, 600 ms, or 800 ms) and then a beep sound was delivered. 1000 ms following the beep sound, participants were presented with the interval estimation scale, which required them to estimate the delay between the key press sound and the beep sound. Participants were told that this delay would be randomly determined in each trial and could not exceed 1000 ms, and they did not receive any further information or training on time estimation. Participants used their right hand to move the mouse cursor along the scale and clicked to indicate their estimation. Inter-trial interval was a 1000 ms blank screen.

In the free choice condition, participants were told that after the star sign was displayed, they should decide for each the trial which key to press and press it at their own pace. The trial-start signal was followed by a blank screen presented for a random delay between 1000 and 2900 ms, after which the star sign was presented and remained on the screen until participants pressed the right or left key. The jittered delay before the star sign was used to avoid routinized key responses and was comparable to the time lag after which an action instruction was given in the human- and robot-instructed conditions. The star sign disappeared with the key press and a blank screen was presented for one of four delays before the beep sound was presented. At the end of each trial, participants indicated their estimation of the delay between their key press and the beep sound.

In the human-instructed condition, the experimenter sat next to the participant where she was able to gaze at both the computer screen and Zora standing behind the separator (see Fig. 1). Participants were told that in this session, they would wait for the experimenter's instruction and press the instructed key (right or left) at a time of their choice. In each trial, the trial-start signal was followed by the star sign. After a random delay between 1000 and 2900 ms, Zora's right or left eye LED briefly turned green, hence cueing the experimenter which instruction (right or left) to give. The rationale behind receiving instruction cues from Zora was to ensure that the experimenter gave randomly determined yet equibalanced number of right and left commands. Participants pressed the instructed key at their own pace and the rest of the trial proceeded as in the free condition.

At the beginning of the robot-instructed block, the experimenter moved the separator and introduced Zora as an autonomous or a non-autonomous robot (see above) and participants were accordingly told that Zora would tell them which key to press. Participants in the autonomous group pressed a key to start the session which triggered Zora to walk down on the desk and stop at a position (fixed vertically in between the participant and the screen while horizontally on the left hand side of the participant, see Fig. 1). First, Zora (looking at the participant) greeted the participant by saying "Hello! I am Zora! Now, I will decide and tell you which key you should press". After that, Zora sat down, looked at the screen and said, "Please press a key to start the session". For this group, Zora also engaged in giving further experimental instructions such as announcing the break time ("Now, you can take a short break and press a key to continue when ready") or signalling the end of the session ("I am done now, thank you, bye!"). For the non-autonomous group, the experimenter moved Zora to the same place at the beginning of the session. Zora did not interact with the participants in this group other than simply speaking out the key press instructions in each trial. Trial procedure in the robot-instructed condition for both groups was the same as the human-instructed condition. During this block of trials, the experimenter remained disengaged with the experiment and seated at the other end of the room.

At the end of each block except the passive condition, participants were asked to rate on the on-screen scale how much control they experienced overall in the preceding block of trials (1: very weak, 6: very strong).

Upon completion of the computer task, participants completed a short questionnaire on which they indicated the degree to which the given adjectives described the robot (see Appendix A). Items in this questionnaire were to quantify the participants' perception of the robot in terms of anthropomorphism, likeability, and perceived intelligence (Bartneck, Kulić, Croft, & Zoghbi, 2009). In addition, participants indicated the degree to which they agreed with two statements (1: strongly disagree, 5: strongly agree) that were aimed to assess the perceived intentionality of the robot ("The robot acted intentionally", Stenzel et al., 2012) and its ability to make its own decisions ("The robot appeared to be have the ability to make its own decisions", van der Woerdt & Haselager, 2017). Finally, participants indicated the number of times they had interacted with a humanoid robot (0–3 or more). Upon completing the questionnaire, participants were debriefed on the goal and experimental manipulations of the study and received their compensation.

2.1.4. Data processing

2.1.4.1. Raw data outlier exclusion Trials in which key press responses were incompliant with the action instruction in the human- and robot-instructed conditions as well as those with interval estimations being three standard deviations away from the mean in each condition were excluded (total excluded: $M = 0.66\% \pm 0.64\%$ of all trials).

2.1.4.2. Participant exclusion Participant exclusion criteria were the proportion of excluded trials being greater than 20% of all trials, failing to follow the experimental instructions, or demonstrating a non-significant trend of increase across the estimations of 200 ms, 400 ms, 600 ms, and 800 ms delays. In order to assess the data against the third criterion, interval estimations and actual delays for each participant were subjected to linear trend analysis (coefficients: −3, −1, 1, 3 for 200 ms, 400 ms, 600 ms, and 800 ms, respectively). None of the participants' data had to be excluded based on these criteria.

2.2. Results

Data analyses were conducted using IBM SPSS 25 software. Significance level was set to 0.05, *post hoc* multiple comparisons were performed using Holm-Bonferroni correction (Holm, 1979), and p values were reported after Holm-Bonferroni procedure. Analysis of Variance (ANOVA) results were reported after Greenhouse-Geisser correction where Mauchly's test of sphericity was violated. Pairwise comparisons were reported with their two-tailed p values unless directional predictions were tested (see Section 1).

2.2.1. Questionnaire items

Mean scores of each questionnaire item for each group are shown in Fig. 2. Shapiro-Wilk tests showed that the data pertaining to the decision-making, intentionality, and previous experience items were not normally distributed ($Ws > 0.8$, $ps < 0.001$), all questionnaire items were thus compared across the two groups by conducting non-parametric Mann-Whitney U tests. The tests showed that participants in the autonomous group ($M = 3.23 \pm 1.33$, $Mdn = 4.00$) believed more strongly compared to the non-autonomous group ($M = 2.37 \pm 1.33$, $Mdn = 2.00$) that the robot could autonomously make its own decisions ($U = 289.50$, $p = .007$, *one-tailed*). All remaining differences in questionnaire items between the two groups were non-significant ($ps > 0.357$).

The relationships among the questionnaire items were examined using Spearman correlation analyses with bootstrapping method with 10.000 iterations. Accordingly, the amount of previous experience with humanoid robots was not related to any item. Anthropomorphism scores, however, were significantly correlated with the scores of perceived intelligence ($\rho = 0.50$, $p < .001$, 95% CI [0.28 0.67]) and decision-making ($\rho = 0.42$, $p = .001$, 95% CI [0.18 0.62]). Perceived intelligence was also correlated with inten-

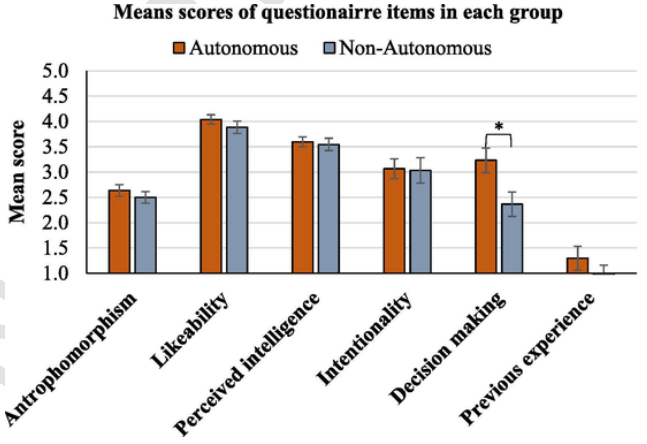

Fig. 2. Mean scores of each questionnaire item for autonomous and non-autonomous groups in Experiment 1. Error bars represent standard error of the mean (*$p < .05$).

tionality ($\rho = 0.47$, $p < .001$, 95% CI [0.20 0.70]) and decision-making ($\rho = 0.46$, $p < .001$, 95% CI [0.20 0.67]). Finally, intentionality and decision-making were significantly correlated ($\rho = 0.28$, $p = .032$, 95% CI [−0.01 0.56]).

2.2.2. Intentional binding

Mean interval estimations for each action-choice condition (passive, free, human-instructed, robot-instructed) for each group of robot-autonomy (autonomous vs. non-autonomous) are shown in Table 1. Intentional binding scores were determined by subtracting the mean interval estimation in each active condition (free, human-instructed, and robot-instructed) from the passive condition in which participants passively judged the interval between two external events (Poonian & Cunnington, 2013). Thus, positive differences between the passive and active conditions would indicate that participants perceived the delay in the passive condition longer compared to the active conditions, in line with the originally reported intentional binding effect (Haggard et al., 2002).

Binding scores (see Fig. 3) were subjected to a 3 × 2 mixed-design repeated measures ANOVA with action-choice (free, human-instructed, robot-instructed) as the within-subjects factor and robot-autonomy (autonomous, non-autonomous) as the between-subjects factor. The test yielded a significant main effect of action-choice ($F(1.76,102.29) = 5.43$, $p = .008$, $\eta_p^2 = 0.09$). The interaction between action-choice and robot-autonomy was not significant ($F < 1$, $p > .8$). *Post hoc* pairwise comparisons showed that binding was significantly stronger in the free condition ($M = 70 \pm 109$) compared to both human-instructed ($M = 37 \pm 101$, $t(59) = 2.79$, $p = .010$, *one-tailed*, $d_z = 0.36$, 95% CI [0.10 0.62]) and robot-instructed ($M = 41 \pm 85$, $t(59) = 2.46$, $p = .017$, *one-tailed*, $d_z = 0.32$, 95% CI [0.06 0.58]) conditions. The difference between human- and robot-instructed conditions was not significant ($t(59) = 0.51$, $p = .610$, $d_z = 0.07$, 95% CI [−0.19 0.32]).

2.2.3. Control ratings

Control ratings were analyzed by a 3 × 2 mixed-design repeated measures ANOVA with action-choice (free, human-instructed, robot-instructed) as the within-subjects factor and robot-autonomy (autonomous vs. non-autonomous) as the between-subjects factor (see Fig. 4). The test revealed a significant main effect of action-choice ($F(1.77,102.85) = 4.05$, $p = .024$, $\eta_p^2 = 0.06$) while action-choice × robot-autonomy interaction was not significant ($F(2,116) = 1.36$, $p = .261$, $\eta_p^2 = 0.02$). *Post hoc* multiple comparisons indicated that participants reported stronger control over the outcome tone in the free condition ($M = 2.80 \pm 1.22$) compared to both human-($M = 2.47 \pm 1.23$, $t(59) = 2.27$, $p = .029$, *one-tailed*, $d_z = 0.29$, 95% CI [0.03 0.55]) and robot-instructed conditions ($M = 2.53 \pm 1.21$, $t(59) = 2.41$, $p = .029$, *one-tailed*, $d_z = 0.31$, 95% CI [0.05 0.57]). The difference between human-instructed and robot-instructed conditions was not significant ($t(59) = 0.52$, $p = .603$, $d_z = 0.07$, 95% CI [-0.19 0.32]).

2.2.4. Relationship between questionnaire items and SoA measures

A critical analysis to the goal of the current study was concerned with the relationship between the SoA measures in the robot-instructed condition and how autonomous the robot was perceived by the participants. To this end, Spearman correlation analyses were conducted, and the results demonstrated that none of the items (i.e., anthropomorphism, likeability, perceived intelligence, intentionality, and decision-making) was significantly correlated with binding ($p > .4$) or control ratings ($p > .08$).

Table 1
Mean interval estimations (collapsed across actual intervals) and standard deviations in each robot-autonomy and action-choice condition in Experiment 1.

Action-choice (within-subjects)	Robot-autonomy (between-subjects)	
	Autonomous	Non-autonomous
Passive	414 ± 89	426 ± 131
Free	343 ± 110	357 ± 135
Human-instructed	370 ± 98	395 ± 127
Robot-instructed	370 ± 113	386 ± 118

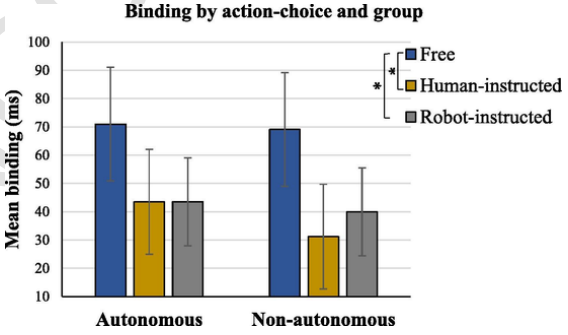

Fig. 3. Mean binding (difference between the interval estimations in passive and active conditions) in each robot-autonomy and action-choice condition in Experiment 1. Error bars represent standard error of the mean (*$p < .05$).

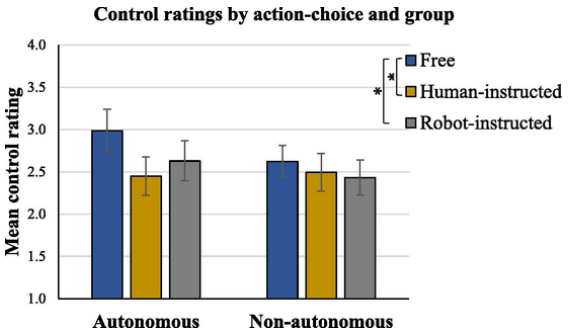

Fig. 4. Mean control ratings in each robot-autonomy and action-choice condition in Experiment 1. Error bars represent standard error of the mean (*$p < .05$).

2.2.5. Relationship between intentional binding and control ratings

Pearson correlation coefficient for each participant was first calculated for the relationship between mean binding and control ratings across the three levels of action-choice (i.e., free, human-instructed, robot-instructed). The resulting coefficients were then tested against 0, which indicated that the correlation between binding and control ratings was not significant ($r_{mean} = 0.04$, $SD = 0.72$, $t(58) = 0.31$, $p = .762$, 95% CI [−0.22 0.29]).

2.3. Discussion

Experiment 1 examined intentional binding and control ratings when actions were freely selected or performed as instructed by a human or a humanoid robot. Importantly, the robot was introduced as an autonomous or a non-autonomous agent to two groups of participants. Furthermore, a questionnaire enabled the investigation of the relationship between the SoA measures and how the robot was perceived in certain human-like features.

To begin with the comparison of the two groups of robot-autonomy, the belief manipulation led to higher ratings of the decision-making ability of the robot by the autonomous compared to the non-autonomous group. This result suggests that the belief manipulation through verbal descriptions could be effective. However, an alternative possibility to the impact of verbal descriptions could be based on the differences in the robot behavior between the two groups. More clearly, the robot behaved more human-like for the autonomous group by interacting with the participants throughout the session while for the non-autonomous group, it remained completely passive and disengaged other than simply uttering action instructions. Although both groups reported similar ratings for the anthropomorphism, the difference in how active and engaging the robot was could have potentially affected the ratings of decision-making ability. This potential confound of robot behavior on perceived decision-making ability is to be eliminated in Experiment 2. Notably, however, belief manipulation regarding the autonomy of the robot did not affect perceived intentionality, and decision-making ability was only weakly correlated with perceived intentionality. This could imply that autonomy and decision-making ability, at least in the present context, was related more strongly to an intellectual capacity than intentionality in the sense of selecting an action with the purpose of achieving a specific outcome. Indeed, perceived intelligence scores were positively correlated with anthropomorphism, decision-making ability, and intentionality. Additionally, anthropomorphism was correlated with perceived decision-making ability, indicating that perceiving an artificial agent endowed with more human-like features is linked to stronger belief in the agent's ability to make autonomous decisions.

Of critical importance to the current study was the finding that perceived autonomy did not influence intentional binding or control ratings. Furthermore, similar binding was observed in human- and robot-instructed actions. However, intentional binding was stronger for freely selected actions compared to when performing human- and robot-instructed actions. This finding supports the results of previous studies reporting stronger binding in freely selected actions as compared to when actions were cued on a computer screen (Barlas & Kopp, 2018; Barlas et al., 2017, 2018) or when instructed by another human (Caspar et al., 2016, 2018).

Control ratings as a direct measure of the SoA displayed a similar trend as intentional binding across the action-choice conditions. More specifically, participants reported stronger control when they could freely select which key to press compared to when they were instructed by either the experimenter or the robot. Although this finding is in line with previous studies demonstrating higher feeling of control ratings in freely selected compared to externally determined actions (Barlas & Kopp, 2018; Barlas et al., 2017, 2018), a limitation in measuring judgment of control Experiment 1 was that participants reported their experience of control as an overall judgment once at the end of each corresponding block. This clearly needs to be improved by obtaining the control ratings in separate blocks of trials.

In summary, Experiment 1 showed that the source of action-choice is an important determinant of subjective experience of control and perceived temporal proximity of actions and their outcomes. Stronger experience of control and binding were observed when actions were freely selected compared to performing actions instructed by either another human or a robot. Importantly, neither binding nor control judgments were affected by the identity or perceived autonomy of the instructor. In this experiment, however, there was no specific relationship between two different actions and the outcomes they produce, as both actions produced the same auditory outcome that did not convey a specific meaning. An intriguing question is then whether the perceived capacity to in-

struct purposeful actions toward a specific outcome could affect how one experiences agency over the outcomes that convey emotional valence. Would individuals, for instance, experience weaker agency for negative outcomes when actions leading these outcomes were purposefully instructed by another person compared to when the instruction was given by an agent ignorant of the possible outcomes? The goal of Experiment 2 was to investigate the SoA in such a context in which participants performed free and instructed actions (as in Experiment 1) that could produce positive, negative, or neutral outcomes.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Experiment 2 aimed at recruiting at least the same sample size as in Experiment 1. Nonetheless, due to the constraints on time and resources, 51 participants in total could be recruited from Bielefeld University (27 females, 5 left-handed, M_{age} = 23.96 years, SD = 4.29 years). Three participants had to be excluded based on the participant exclusion criteria (see Section 3.1.4) and thus data of 48 participants were subjected to the analyses (26 females, 5 left-handed, M_{age} = 23.81 years, SD = 4.30 years). Participants were randomly assigned to one of the autonomous (n = 24, 13 females, 3 left-handed, M_{age} = 23.71 years, SD = 4.36 years) and non-autonomous groups (n = 24, 13 females, 2 left-handed, M_{age} = 23.92 years, SD = 4.33 years). All participants had normal or corrected-to-normal vision and had no hearing problems. Participants gave their written consent prior to the study and received monetary compensation (10 Euros) in exchange for their participation. The study was conducted in accordance with the ethical guidelines of the Declaration of Helsinki and was approved by the Research Ethics Board of Bielefeld University.

3.1.2. Materials and experimental setup

The same stimuli and experimental setup as in Experiment 1 were used except for the following amendments. First, outcome-valence was manipulated using face images with neutral, happy, and disgusted expressions. Accordingly, face stimuli consisted of 72 images (6 images for each female/male subject) selected among the NimStim collection of face stimuli (Tottenham et al., 2009). These face images displayed neutral, positive (happy), and negative (disgusted) expressions (see Appendix B for a complete list of face stimuli). All images were pre-processed to cut out hair of each subject as it was irrelevant to the expressions and could potentially be distracting. Each resulting image was at a dimension of 506 × 617 pixels. Second, all stimuli were presented against a black-background and action-outcome delays consisted of 200, 500, and 800 ms. And third, response key labels were placed on "v" and "n" keys for left and right responses, respectively. This change was to avoid having to look at the keyboard as participants were instructed to monitor the computer screen throughout the experiment. Accordingly, they placed their left index finger between the two keys at the beginning of each trial and simply moved their finger onto the right or left key to perform the key presses.

3.1.3. Procedure

The duration of the study was approximately 90 min. Experiment 2 followed the setup and procedure in Experiment 1 with the following amendments. Participants in the autonomous group were told that Zora was an autonomous robot capable of making its own decisions, and when Zora told them which key to press, it *knew* how pressing that key would change the expression on the face. This description was intended to convey the message that the robot gave action instructions with a specific purpose. In accordance with this description, Zora greeted the participants in this group with "Hello! I am Zora! Now, I will decide and tell you which key you should press. When I say 'right' or 'left', I know how that will change the expression of the face". The non-autonomous group was told that Zora's key press instructions were pre-programmed in a random order and Zora would simply tell them which key to press. Zora greeted the participants in this group with "Hello! I am Zora! Now, I will tell you which key you should press". Otherwise the robot behavior (e.g., walking, greeting the participants) was the same in both autonomous and non-autonomous conditions, which was to eliminate the potential confound of robot behavior on perceived robot-autonomy (see Section 3.3). Finally, in the human-instructed condition, all participants were told that the experimenter knew how her action instruction could change the target face.

Participants first completed the interval estimation task followed by the judgment of control task, and finally completed the post-experiment questionnaire as in Experiment 1. Interval estimation task consisted of 288 experimental trials in total and was presented in four blocks of action-choice conditions (i.e., passive, free selection, human-instructed, robot-instructed). The order of the action-choice blocks was counterbalanced across participants. All facial expressions (neutral, happy, disgusted) and action-outcome intervals (200 ms, 500 ms, 800 ms) were presented in a random order within each block. Judgment of control task included all but the passive block with 36 trials each (108 in total). Since the effect of action-outcome interval on subjective judgment of control was not of interest here, action-outcome interval was fixed at 200 ms. At the beginning of each block, participants completed 10 additional practice trials in both interval estimation and judgment of control tasks. Additionally, in order to make sure participants paid attention to the facial stimuli, 5 randomly determined trials in each block served as catch trials, at the end of which participants were asked to identify the expression on the last face they were presented with. The question was presented on the screen ("What was the emotion on the last face image?") with three response choices ("neutral", "happy", or "disgusted") and participants indicated their response with a mouse-click on one of three options. They received feedback immediately after their response (i.e., "That was correct!" or "That was wrong, please pay attention to the face images").

Each trial in the passive condition of the interval estimation task started with a fixation cross presented for 500 ms. This was followed by the presentation of a target face image (always with a neutral expression) that remained on the screen for a duration jittered between 1000 and 2900 ms, after which the face image disappeared simultaneously with a key press sound. A blank screen was then presented for a random delay (200 ms, 500 ms, 800 ms) and followed by the beep sound simultaneously presented with the outcome face image. The outcome face image portrayed the same subject in the target face image, albeit with another expression of a neutral, happy, or disgusted affect (see Fig. 5). The outcome face remained on the screen for 1000 ms. After a 500 ms delay following the image offset, the interval estimation scale was presented, and participants indicated their estimate of the delay between the key press sound and the onset of the outcome image. As in Experiment 1, they were told that this delay would be randomly determined in each trial and could not exceed 1000 ms. Inter-trial interval was a 1000 ms blank screen.

In the free selection condition, participants were told that could freely choose between the right and the left key and press it at their own pace after the target face image was presented. In this condition, the fixation cross was presented for 1000–2900 ms and was followed by the presentation of the target face. As in Experiment 1, this delay presented before the target face served to avoid routinized key responses and more importantly, corresponded to the time lag after which an action instruction was given in the human- and robot-instructed conditions. The target face disappeared with the key press and after one of three delays (200 ms, 500 ms, 800 ms) and the outcome face image was presented simultaneously with the beep sound. The expression of the outcome face was randomly chosen for each key press and therefore, outcomes were not contingent on the key press actions. At the end of each trial, participants indicated their estimation of the delay between their key press and the onset of the outcome face image.

In the human- and robot-instructed conditions, the target face was presented after a 500 ms display of a fixation cross. Depending on the condition, the robot or the experimenter gave the action instruction (i.e., "right" or "left)" between 1000 and 2900 ms after the onset of the target face. Participants pressed the instructed key at a time of their choice, which replaced the target image with the outcome face presented simultaneously with the beep sound after a random delay (200 ms, 500 ms, 800 ms). The outcome image was presented for 1000 ms and participants then reported their estimation of the delay between their key press the onset of the outcome face image.

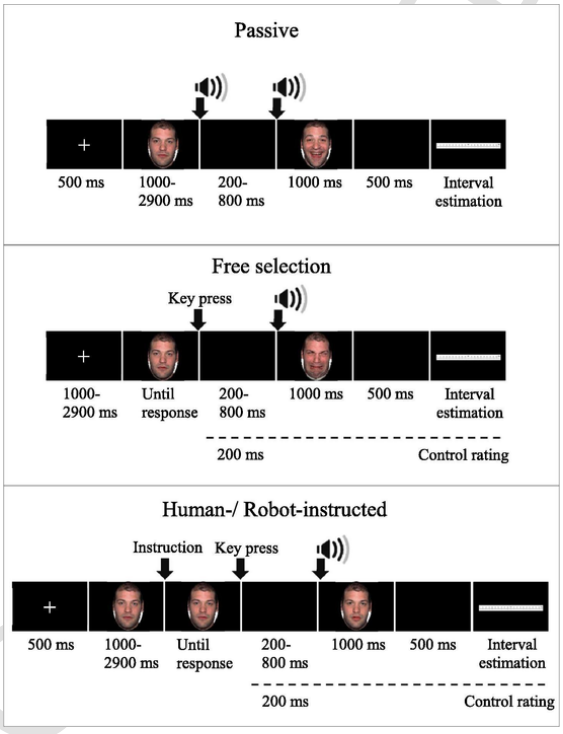

Fig. 5. Schematic illustration of the trial procedure in each condition in Experiment 2. Participants completed the interval estimation and judgment of control tasks in that order. In the passive condition, participants estimated the delay between the key press click sound and the onset of the second face image. In the free condition, they freely chose between the right and the left key while in the human and robot-instructed conditions, they pressed the instructed key. At the end of each trial in these conditions, they were asked to estimate the delay between their key press and the onset of the outcome face image. Upon completing the four blocks of interval estimation task, participants performed three blocks of judgment of control task in which for each trial, they indicated how much control they experienced over the change of the facial expression after their key press (1: very weak; 6: very strong). Additionally, as in Experiment 1, participants completed a post-experiment questionnaire that assessed the robot in terms of anthropomorphology, likeability, perceived intelligence, and whether it appeared to give intentional action instructions and capable of making its own decisions.

Interval estimation task was followed by the judgment of control task which consisted of free, human- and robot-instructed conditions. Trial procedure in the task was the same, except that the action-outcome interval was fixed at 200 ms and participants reported how strong was their experienced control over the change of the facial expression in each trial (1: very weak, 6: very strong).

Upon completion of the computer task, participants completed the post-experiment questionnaire (see Appendix A). Finally, participants were debriefed on the purpose of the study and the experimental manipulations and received their compensation.

3.1.4. Data processing

3.1.4.1. Raw data outlier exclusion Trials with incompliant key responses (in human – and robot-instructed conditions), incorrectly responded catch trials, or those with interval estimations being three standard deviations away from the mean were excluded (interval estimation task: $M = 1.43\% \pm 1.13\%$; judgment of control task: $M = 0.71\% \pm 1.43\%$).

3.1.4.2. Participant exclusion Participant exclusion criteria were the proportion of excluded trials being greater than 20% of all trials, failing to follow the experimental instructions, and demonstrating a non-significant linear trend across the estimations of 200 ms, 500 ms, and 800 ms delays. Two participants were excluded based on the last criterion. Due to a technical problem, one participant's data were incomplete and thus excluded from the analyses.

3.2. Results

As in Experiment 1, data analyses were conducted using IBM SPSS 25 software. Significance level was set to 0.05, *post hoc* multiple comparisons were performed using Holm-Bonferroni correction (Holm, 1979), and *p* values were reported after Holm-Bonferroni procedure. Analysis of Variance (ANOVA) results were reported after Greenhouse-Geisser correction where Mauchly's test of sphericity was violated. Pairwise comparisons were reported with their two-tailed *p* values unless directional predictions were tested (see Section 1).

3.2.1. Questionnaire items

Mean scores of each questionnaire item for each group are shown in Fig. 6. Shapiro-Wilk tests showed that data pertaining to the likeability ($p = .012$), decision-making, and intentionality items were not normally distributed ($p < .001$), Mann-Whitney *U* tests were thus conducted to compare the scores between the two groups. The tests revealed that perceived ability of decision-making was higher in the autonomous group ($M = 2.96 \pm 1.23$, $Mdn = 3.00$) than the non-autonomous group ($M = 2.25 \pm 1.19$, $Mdn = 2.00$, $U = 198$, $p = .028$, *one-tailed*). Concerning the remaining items of the questionnaire, as in Experiment 1, no significant differences were observed between the two groups ($ps > 0.092$).

In order to assess the relationships among the questionnaire items, Spearman correlation analyses were conducted. As in Experiment 1, the amount of previous experience with humanoid robots was not related to any other item. Anthropomorphism scores were significantly correlated with the scores of likeability ($\rho = 0.39$, $p = .006$, 95% CI [0.12 0.62]), perceived intelligence ($\rho = 0.39$, $p = .006$, 95% CI [0.14 0.60]), and decision-making ($\rho = 0.35$, $p = .013$, 95% CI [0.09 0.59]). Perceived intelligence was correlated with decision-making ($r = 0.30$, $p = .040$, 95% CI [0.00 0.57]) and likeability ($\rho = 0.29$, $p = .044$, 95% CI [0.02 0.53]). Finally, decision-making was correlated with intentionality ($r = 0.35$, $p = .015$, 95% CI [0.02 0.62]).

3.2.2. Intentional binding

Mean interval estimations in each action-choice (passive, free, human-instructed, robot-instructed) and outcome-valence (neutral, positive, negative) conditions for each group of robot-autonomy (autonomous vs. non-autonomous) are shown in Table 2. As in Experiment 1, binding scores were calculated by subtracting the mean interval estimations in each action-choice and outcome-valence condition from the corresponding valence level in the passive condition (see Fig. 7).

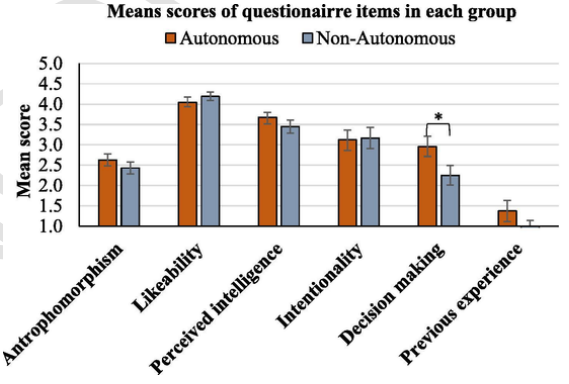

Fig. 6. Mean scores of each questionnaire item for autonomous and non-autonomous groups in Experiment 2. Error bars represent standard error of the mean (*$p < .05$).

Table 2
Mean interval estimations (collapsed across actual delays) and standard deviations in each action-choice and outcome-valence condition for autonomous and non-autonomous groups in Experiment 2.

Action-choice (within-subjects)	Robot-autonomy (between-subjects)					
	Autonomous			Non-autonomous		
	Neutral	Positive	Negative	Neutral	Positive	Negative
Passive	389 ± 111	397 ± 123	397 ± 119	396 ± 123	416 ± 125	422 ± 114
Free	335 ± 117	338 ± 115	351 ± 115	356 ± 88	372 ± 86	379 ± 88
Human-instructed	369 ± 105	382 ± 91	394 ± 95	395 ± 100	392 ± 103	402 ± 102
Robot-instructed	338 ± 102	371 ± 118	381 ± 123	368 ± 98	391 ± 109	383 ± 101

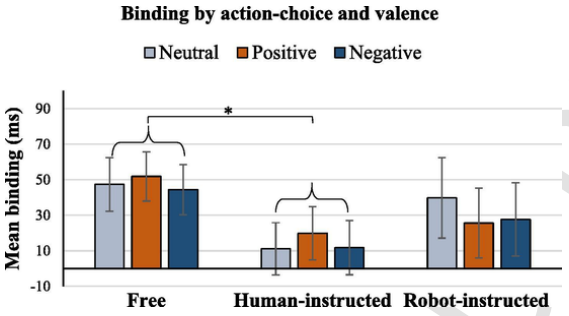

Fig. 7. Mean binding (difference between the interval estimations in passive and active conditions) in each action-choice and outcome-valence condition plotted overall for both groups in Experiment 2. Error bars represent standard error of the mean (*$p < .05$).

Binding scores were subjected to a 3 × 3 × 2 mixed-design repeated measures ANOVA with action-choice (free, human-instructed, robot-instructed), valence (neutral, positive, negative) as within-subjects factors and robot-autonomy (autonomous, non-autonomous) as the between-subjects factor. The test revealed a main effect of action-choice ($F(2,92) = 5.45$, $p = .006$, $\eta_p^2 = 0.11$). *Post hoc* multiple comparisons among the action-choice conditions suggested that, as predicted, binding was stronger in the free condition ($M = 48 \pm 92$) compared to the human-instructed condition ($M = 14 \pm 96$, $t(47) = 3.21$, $p = .003$, *one-tailed*, $d_z = 0.46$, 95% CI [0.16 0.76]). Differences between the free and robot-instructed ($M = 31 \pm 97$, $t(47) = 1.55$, $p = .064$, *one-tailed*, $d_z = 0.22$, 95% CI [-0.06 0.51]) as well as between the human-instructed and robot-instructed conditions were not significant ($t(47) = 1.90$, $p = .128$, $d_z = 0.27$, 95% CI [-0.02 0.56]). The main effect of valence was not significant ($F(2,92) = 0.30$, $p = .744$, $\eta_p^2 = 0.01$), nor were the action-choice × valence, action-choice × robot-autonomy, and action-choice × valence × robot-autonomy interactions ($Fs < 1.9$, $ps > 0.1$).

3.2.3. Control ratings

Judgment of control ratings were analyzed by a 3 × 3 × 2 mixed-design repeated measures ANOVA with action-choice (free, human-instructed, robot-instructed) and outcome-valence (neutral, positive, negative) as the within-subjects factors and robot-autonomy (autonomous vs. non-autonomous) as the between-subjects factor. The test revealed a significant main effect of action-choice ($F(1.73,79.66) = 4.38$, $p = .020$, $\eta_p^2 = 0.09$) and a main effect of valence ($F(1.66,76.60) = 9.07$, $p = .001$, $\eta_p^2 = 0.16$). All remaining effects were non-significant ($Fs < 1.8$, $ps > 0.1$).

Post hoc multiple comparisons indicated that participants reported stronger control over the outcome in the free condition ($M = 2.73 \pm 1.03$) compared to human-instructed ($M = 2.49 \pm 1.04$, $t(47) = 2.79$, $p = .012$, *one-tailed*, $d_z = 0.40$, 95% CI [0.11 0.70]) and robot-instructed ($M = 2.51 \pm 1.05$, $t(47) = 2.11$, $p = .040$, *one-tailed*, $d_z = 0.30$, 95% CI [0.01 0.59]) conditions (see Fig. 8). The difference between human-instructed and robot-instructed conditions was not significant ($t(47) = 0.28$, $p = .780$, $d_z = 0.04$, 95% CI [−0.24 0.32]). With respect to the effect of outcome-valence, control ratings were higher in the positive ($M = 2.91 \pm 1.27$) compared to both neutral ($M = 2.26 \pm 0.99$, $t(47) = 4.09$, $p < .001$, *one-tailed*, $d_z = 0.59$, 95% CI [0.28 0.89])) and negative ($M = 2.56 \pm 1.16$, $t(47) = 3.05$, $p = .004$, *one-tailed*, $d_z = 0.44$, 95% CI [0.14 0.73]) conditions. In contrast to the prediction that negative outcomes would yield lower control ratings compared to the neutral outcomes, control ratings were higher for negative ($M = 2.56 \pm 1.16$) than neutral ($M = 2.26 \pm 0.99$) outcomes. The difference, however, was not significant ($t(47) = 1.71$, $p = .093$, $d_z = 0.25$, 95% CI [−0.04 0.53]).

3.2.4. Relationship between questionnaire items and SoA measures

As in Experiment 1, Spearman correlation analyses were conducted to examine the relationship between the items of the post-experiment questionnaire and binding and control ratings. Again, binding in the robot-instructed condition was not correlated with any of the item ($p > .2$). Mean control ratings in the robot-instructed condition, however, was positively correlated with anthropo-

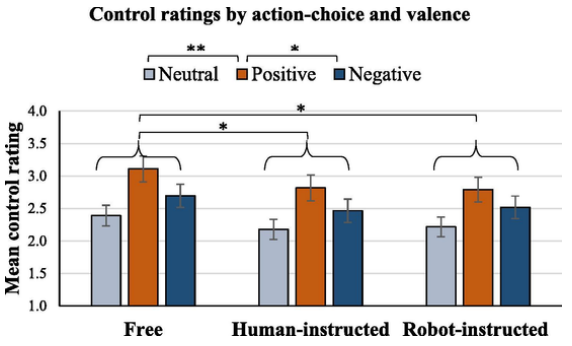

Fig. 8. Mean control ratings in each action-choice and outcome-valence condition plotted overall for both groups in Experiment 2. Error bars represent standard error of the mean (*$p < .05$; **$p < .001$).

morphism ($\rho = 0.45$, $p = .001$, 95% CI [0.19 0.66]), likeability ($\rho = 0.44$, $p = .002$, 95% CI [0.16 0.61]), perceived intelligence ($\rho = 0.29$, $p = .043$, 95% CI [0.01 0.56]), intentionality ($\rho = 0.29$, $p = .043$, 95% CI [−0.01 0.58]), and decision-making ability ($\rho = 0.35$, $p = .016$, 95% CI [0.05 0.60]).

3.2.5. Relationship between intentional binding and control ratings

As in Experiment 1, Pearson correlation coefficients of the relationship between binding and control ratings were calculated for each participant across the levels of action-choice and outcome-valence, and then were tested against 0. The tests did not reveal any significant correlations between binding and control ratings across the levels of action-choice ($r_{mean} = 0.19$, $SD = 0.73$, $t(46) = 1.31$, $p = .098$, 95% CI [-0.01 0.45]) and outcome-valence ($r_{mean} = 0.09$, $SD = 0.68$, $t(46) = 0.61$, $p = .543$, 95% CI [−0.20 0.37]).

3.3. Discussion

The results of Experiment 2 regarding the questionnaire items were mainly in line with the findings Experiment 1. Accordingly, participants in the autonomous group reported that the robot more strongly appeared capable of making its own decisions compared to the non-autonomous group. It was previously noted that one alternative explanation for the difference found between the two groups in Experiment 1 could be merely driven by the lack of engagement of the robot for the non-autonomous group. Here in Experiment 2, the robot behavior was the same for both groups and thus, reports of perceived decision-making ability of the robot suggest that the experimental (belief) manipulation through verbal descriptions of the robot-autonomy alone could be effective. Secondly, examination of the relationships among the questionnaire items partly confirmed the results of Experiment 1 in that anthropomorphism was positively correlated with decision-making and perceived intelligence, and decision-making was correlated with perceived intelligence. Experiment 2 additionally showed that likeability was positively correlated with anthropomorphism, and perceived intelligence and decision -making were correlated with intentionality.

Of more interest was the results concerning the effects of belief in robot-autonomy, action-choice, and outcome-valence on the SoA measures. Confirming the results of Experiment 1, perceived robot-autonomy did not affect intentional binding or control ratings when participants performed robot-instructed actions. However, stronger binding was observed when participants were free to choose their actions compared to when they performed human-instructed actions while binding in the robot-instructed condition was not significantly different from either free or human-instructed conditions. The results regarding the effect of action-choice on control ratings confirmed the findings of Experiment 1, showing that stronger control was experienced over the outcomes produced by freely selected compared to instructed actions.

It is important to note that the effect of action-choice on binding and control ratings was independent of the valence of the outcomes. Furthermore, binding and control ratings displayed distinct responses to outcome-valence. More clearly, binding was found to be unaffected by the valence of outcomes while control ratings were higher when participants' actions changed the neutral expression on the face stimuli to a positive expression compared to when the expression remained neutral or changed to a negative affect. Thus, positive outcomes could have an enhancing effect on explicit experience of control while negative outcomes were experienced as comparable to neutral outcomes.

Recent work on the effect of outcome-valence on intentional binding suggested that this effect might be confined with the predictability of the valence of outcomes (Yoshie & Haggard, 2017). Indeed, among previous studies, those rendered the outcomes contingent and predictable on the corresponding actions reported stronger binding for positive compared to negative outcomes (Barlas & Obhi, 2014; Takahata et al., 2012; Yoshie & Haggard, 2013) while no effect of valence was found in studies in which the valence of outcomes was not associated with the actions that produced them (Barlas et al., 2017, 2018). As the right and left key presses in the current study randomly produced neutral, positive, or negative outcomes, the finding that valence did not affect binding might comply with the predictability account of the effect of outcome-valence on binding. Yet, further research is required to address the divergent findings that demonstrated a null effect of valence on binding even when the valence of outcomes were predictable (Moreton, Callan, & Hughes, 2017).

An interesting finding was that control ratings were positively related with all questionnaire items that assessed how human-like the robot was perceived. More clearly, the stronger the robot was perceived as anthropomorphic, likeable, intelligent, intentional, and able to autonomously make its decisions, the stronger control over the action-outcomes was experienced when performing robot-instructed actions. Although receiving action instructions from a human vs. a robot did not affect the subjective experience of control, this finding might have important implications for designing robots that are to guide human actions.

4. General discussion

Previous research examining the SoA in joint actions with artificial agents and humans reported a weakened experience of control when the interaction partner was an artificial agent compared to when it was another person (Obhi & Hall, 2011b; Sahaï et al., 2019). The distinct experience of agency when co-acting with humans vs. artificial agents was ascribed to the inability to co-represent the actions and motor plans of artificial agents (Obhi & Hall, 2011b; Wohlschläger et al., 2003), and that predictive mechanisms as described in the Comparator Model may not function comparably for these two different action partners (Sahaï et al., 2017). However, other studies demonstrated that actions of a humanoid robot could be co-represented depending on the perceived intentionality (Stenzel et al., 2012) and more strongly on the perceived agency (Stenzel et al., 2014). Additionally, observing actions of both a human and an anthropomorphic robotic hand could yield similar effects on intentional binding for one's own actions (Khalighinejad, Bahrami, et al., 2016). These results, overall, suggest that human-like features of artificial agents as well as higher level beliefs about their agentic capacity could play a critical role in how people experience agency in joint actions with these agents. In two experiments, the present study examined how SoA—measured by intentional binding and explicit control ratings—would be influenced when actions are determined by another human vs. a humanoid robot as compared to when freely selected. To also investigate whether perceived autonomy of the robot could influence the SoA in robot-instructed actions, both experiments administered a belief manipulation that involved autonomous vs. non-autonomous descriptions of the robot. In Experiment 1, each alternative action produced the same auditory tone while in Experiment 2, the outcomes produced by two actions conveyed neutral, positive, or negative valence.

4.1. Human- vs. robot-instructed actions

The main finding of two experiments was that belief in robot-autonomy did not affect intentional binding or how participants experienced control when performing robot-instructed actions. This result was observed despite that participants who were given an autonomous description of robot more strongly agreed on the autonomous decision-making ability of the robot compared to the group receiving a non-autonomous description. However, these descriptions did not induce any difference between the two groups regarding the intentionality of the robot. It thus seems that the belief manipulation could only affect the perceived decision-making capacity, but not perceived intentionality of the robot. In this regard, the current study failed to replicate the effect of belief manipulation on perceived intentionality as demonstrated by Stenzel et al. (2012), although the robot descriptions were similar between the current study and the study by Stenzel et al. (2012). Notwithstanding with how effective the belief manipulation was, however, another important finding of both experiments was that participants experienced comparable intentional binding and subjective control when they performed human- and robot-instructed actions. Notably, indistinguishable SoA between human and robot-instructed actions was found in both experiments regardless of the value or relevance of the outcomes produced by the instructed actions. These findings are partly in line with a previous finding that no difference in intentional binding was exhibited in one's actions when observing human and robot produced actions (Khalighinejad, Bahrami, et al., 2016). Additionally, the current findings support the previous studies that reported similar Social Simon effect when cooperating with another person and a humanoid robot (Stenzel et al., 2012, 2014), although their finding hinged on perceived intentionality (Stenzel et al., 2012) and more strongly on perceived agency of the robot (Stenzel et al., 2014). However, it should be noted that the current study established a remarkably different methodology in that participants simply received action instructions from the *other agent* rather than observing their actions or acting in cooperation. Correspondingly, the co-representation of actions in the present context was confined to what action was externally selected and instructed by the experimenter or the humanoid robot. Within the boundaries of the current study, therefore, it can be argued that action selection determined by another human and a humanoid robot could be similarly (co-) represented or incorporated into the motor control mechanisms that were to accomplish the externally determined actions. The similarity between how human and robot generated action selection were represented could, as a result, produce comparable intentional binding and explicit control over the action-outcomes between human- and robot-instructed conditions.

4.2. Free vs. externally determined actions

Another important finding was that irrespective of the identity of the instructor, experience of control and intentional binding were stronger when participants freely selected their actions compared to when they performed instructed actions. The role of endogenous processing of action selection in SoA has been emphasized by several studies (Barlas & Obhi, 2013; Barlas et al., 2018; Caspar et al., 2016, 2018). Caspar et al. (2016), for instance, showed that binding was stronger when participants freely chose their action that could harm or not their co-participant compared to when they produced these actions as instructed by the experimenter. Additionally, Barlas et al. (2018) found that both intentional binding and self-reports of experienced control were stronger when

participants could freely choose among four key press alternatives compared to when they were instructed to perform the key press specified on a computer screen. Several neuroimaging and brain stimulation studies have also highlighted the differences in underlying neural structures between internally generated and externally determined actions. These studies showed that free action selection in contrast to instructed actions is associated with increased activity in dorsolateral prefrontal cortex (DLPFC), inferior parietal lobe (IPL), rostral cingulate zone (RCZ), and supplementary motor area (SMA) (Cunnington, Windischberger, Deecke, & Moser, 2002; Filevich et al., 2013; Lau, Rogers, & Passingham, 2006; Lau, Rogers, Haggard, & Passingham, 2004; Waszak et al., 2005). Among these areas, SMA and DLPFC were also found linked to the intentional binding effect. Kühn et al. (2012), for instance, demonstrated that activity in the SMA was positively correlated with the strength of intentional binding, and disrupting pre-SMA by theta-burst transcranial magnetic stimulation was shown to reduce the intentional binding effect (Moore, Ruge, Wenke, Rothwell, & Haggard, 2010). Additionally, anodal stimulation of DLPFC with transcranial direct current stimulation was found to enhance temporal binding of actions and outcomes when the actions were freely chosen (Khalighinejad, Di Costa, & Haggard, 2016). Together, these findings support the notion that free choice and endogenous processing of actions take a critical role in determining the strength of intentional binding as an implicit index of the SoA.

Subjective judgments of control have also been shown to involve several brain regions including temporoparietal junction (TPJ), IPL, SMA, anterior cingulate cortex (ACC) and DLPFC (David, Newen, & Vogeley, 2008; Fukushima, Goto, Maeda, Kato, & Umeda, 2013; Sperduti, Delaveau, Fossati, & Nadel, 2011). Greater activation in these areas when action selection is internally generated could provide one potential explanation for the current finding that explicit control ratings were higher over the outcomes produced by free compared to human- and robot-instructed actions. Alternatively, previous views on the link between autonomy and SoA suggested that free action selection could bolster one's sense of autonomy (Schwartz, 2012), and people may tend to feel stronger control when their actions are based on self-generated decisions and intentions (Haggard, 2008; Sebanz & Lackner, 2007). The contrast between free action selection and being instructed by external sources could therefore enhance how strongly one experiences in control of the outcomes. Interestingly, the current study demonstrated this effect both when the action-outcomes were fixed for both actions (Experiment 1) and when outcomes conveyed a specific valence (Experiment 2). Thus, who determines what action to take prior to movement seems to cast a strong impact on explicit judgments of agency regardless of the changes these actions produce.

4.3. The effect of valence on intentional binding and explicit control ratings

In Experiment 2, key presses could randomly produce a neutral, positive, or a negative expression on the face stimuli. The valence of these outcomes, however, did not influence intentional binding. To reiterate, studies investigating the effect of valence on binding has confronted controversial findings. On the one hand, earlier studies have reported different effects of outcome-valence including the finding of reduced binding with negative compared to positive or neutral outcomes (Yoshie & Haggard, 2013) and enhanced binding with outcomes that are linked to positive monetary gains (Takahata et al., 2012). On the other hand, a direct replication of Yoshie and Haggard (2013)'s study failed to confirm their findings (Moreton et al., 2017), and other studies in which the valence of outcomes was unpredictable reported that binding was immune to the valence of outcomes (Barlas et al., 2017, 2018). One view in response to these seemingly controversial findings suggested that modulation of intentional binding by emotional outcome-valence could depend on the predictability of the valence of outcomes (Yoshie & Haggard, 2017). This could correspondingly indicate that binding may not be modulated by the postdictive information pertaining to the emotional component of action-outcomes and requires a reliable predictive model of the outcome valence (Yoshie & Haggard, 2017). It should not be conceived, however, that postdictive information cannot alter intentional binding. Several previous studies have shown that when the predictive information is not available regarding the occurrence of an outcome (Moore & Haggard, 2008) or the congruency between actions and outcomes (Barlas & Kopp, 2018; Ebert & Wegner, 2010), postdictive evidence informing the outcome occurrence or congruency could alter the intentional binding effect. Clearly, more research is required to unveil the precise mechanisms of how predictive and postdictive cues are integrated depending on the context (Moore & Fletcher, 2012).

In contrast to binding, explicit control judgments seem to rely on the postdictive information regarding the valence of outcomes when this information is not predictable. Indeed, the results in Experiment 2 showed that stronger control was experienced when outcomes conveyed a positive compared to neutral and negative valence. The tendency to attribute positive as opposed to negative events to one's actions complies with the well-known notion of self-serving bias (Duval & Silvia, 2002; Miller & Ross, 1975; Taylor & Brown, 1994). In this vein, the current findings strengthen the proposed link between SoA and self-serving bias (Barlas & Obhi, 2014; Yoshie & Haggard, 2013) in the context of unpredictable outcomes produced by free and instructed actions. The effect of outcome-valence on explicit control ratings also complies with the previous studies that reported stronger control ratings over pleasant compared to unpleasant auditory outcomes (Barlas et al., 2017, 2018). However, these studies did not include a neutral condition and therefore, it was unclear whether the effect of valence was in the direction of enhancement by positive outcomes or attenuation by negative outcomes. In the current study, neutral and negative outcomes were not experienced distinctly by the participants. Thus, it appears that the effect of valence on control ratings was more likely to reflect enhancement of experienced control by positive outcomes.

4.4. Relationship between perceived robot features and SoA measures

An interesting finding was that subjective judgments of experienced control were positively correlated with how human-like some features of the robot were perceived. More specifically, the stronger participants perceived the robot as anthropomorphic, likeable, intelligent, and with the capacity of decision-making and generating purposeful action instructions, the stronger they experienced in control of the outcomes in the robot-instructed condition. Although similar degree of control was experienced between human- and robot- instructed actions, this finding sparks the question whether SoA when receiving action guidance from different artificial agents could be modulated by the anthropomorphic features of these agents. Future studies should investigate this by contrasting artificial agents with varying degrees of anthropomorphic features (e.g., computers, humanoid and non-humanoid robots, etc.). For the moment, the observed relationship between explicit control ratings and human-like features of the robot could provide some input into design of the robots that are to assist and guide human actions.

4.5. Relationship between intentional binding and explicit control ratings

A relatively long-standing question in SoA research is whether and how implicit and explicit measures are related. Yet, relevant studies probing this question have provided rather mixed findings. Dewey and Knoblich (2014), for instance, used the clock paradigm (see Haggard et al., 2002) and additionally measured sensory attenuation of the auditory outcomes and subjective judgments of control over these outcomes. The authors found that neither of these measures was correlated with another. Conversely, in the study by Berberian et al. (2012) described before, intentional binding and control ratings were positively correlated across the levels of autonomy of the flight simulation system. Finally, a recent study by Imaizumi and Tanno (2019) examined inter-individual and intra-individual correlations between intentional binding and agency ratings while manipulating the action-outcome intervals for auditory and visual outcomes. The authors found intra-individual correlations between intentional binding and agency ratings while inter-individual correlations were observed only for the auditory outcomes. The current study failed to find a relationship between intentional binding and control ratings. Although an in-depth consideration of the methodological differences among the previous and current studies is beyond the scope of this paper, the present findings—specifically with respect to the differential effect of outcome valence on intentional binding and control ratings—highlight the notion that these implicit and explicit measures of the SoA might rely on distinct mechanisms (Dewey & Knoblich, 2014; Ebert & Wegner, 2010; Wen et al., 2015).

4.6. Limitations and future directions

One limitation of the current study is that the experimental condition in which action instructions were given by a non-humanoid robot or another artificial agent with less human-like features was not included. Although such an additional condition would have provided more insight into how physical characteristics of artificial agents modulate one's SoA, it would as well render the experiment taking much longer. Given the present study found a positive relationship between explicit control ratings and perceived human-like features of the robot, future work should directly compare SoA when actions are guided by artificial agents that vary in their human-like features. Another limitation was that Experiment 2 could not recruit the number of participants as planned based on Experiment 1 (48 vs. 60), which could have led to the smaller difference found in binding between free and robot-instructed actions compared to Experiment 1. Thus, Experiment 2 might require to be replicated to ensure the reliability of the current results.

A critical aspect of the current study that could be considered as a limitation is that outcomes in Experiment 2 were non-contingent on the actions. Although this might have undermined the notion of purposefulness of the actions and could contribute to the comparable SoA found between human- and robot-instructed actions, it enabled the isolation of the effect of *who* determined the action selection (i.e., internally generated vs. externally given by another human or a humanoid robot). Furthermore, the robot in the autonomous group was introduced as if it knew how the face stimuli would change after the instructed action and at least for the present design, this manipulation could be strengthened when outcomes were unpredictable to the participants. A follow-up study could use a different belief manipulation strategy and include three action alternatives that are associated with neutral, positive, and negative outcomes and complementary to the current findings, this would enable the examination of whether predictability of outcomes would demonstrate a different picture in human- vs. robot-instructed actions.

An interesting question for future studies to investigate would be how one's SoA and feeling of responsibility would be altered when one gives action instructions to an artificial agent as compared to another person. A recent study reported a reduced binding effect both when participants instructed the actions of another person and performed instructed actions (Caspar et al., 2018). Compared to instructing other persons, however, would one feel more in control and accountable for the consequences when they determine what a machine could do or not?

5. Conclusions

With the perpetual advance of technology, artificial agents such as humanoid robots have been more frequently involved in daily lives. In two experiments, the current study provides preliminary steps towards understanding how SoA, a fundamental aspect

of human actions, will alter when human actions are guided by these agents. The main results indicate that individuals may experience similar SoA when they perform actions guided by another human or humanoid robot, independent of how human-like the robot is perceived and the valence of action-outcomes. In addition, the finding that free choice actions yielded stronger binding and experience of control over the outcomes as compared to instructed actions, highlights the critical role of endogenous processing of action selection on the SoA. Finally, a positive correlation found between explicit control ratings and how human-like the robot was perceived could feed into the design of artificial systems. Future studies with other artificial agents of varying anthropomorphic features could provide further insight to our understanding of how interactions with technology could alter one's SoA.

Funding

This work was supported by the Cluster of Excellence Cognitive Interaction Technology "CITEC" (EXC 277) at Bielefeld University, funded by the German Research Foundation (DFG).

Declaration of Competing Interest

The author declares that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

The author would like to thank Stefan Kopp for his support throughout; Thorsten Schodde, Martin Wiechmann, and Cognitive Systems Engineering group for their technical support and providing the NAO robot; Hendrik Buschmeier and Farina Freigang for their help with photographing the experimental setup; and Stanley Indelicato for processing the face stimuli.

Appendix A. Post-experiment questionnaire-Experiment 1&2

Q1) Please rate your impression of the robot on these scales:

1	Fake	1	2	3	4	5	Natural
2	Machine-like	1	2	3	4	5	Human-like
3	Unconscious	1	2	3	4	5	Conscious
4	Artificial	1	2	3	4	5	Lifelike
5	Moving rigidly	1	2	3	4	5	Moving elegantly
6	Dislike	1	2	3	4	5	Like
7	Unfriendly	1	2	3	4	5	Friendly
8	Unkind	1	2	3	4	5	Kind
9	Unpleasant	1	2	3	4	5	Pleasant
10	Awful	1	2	3	4	5	Nice
11	Incompetent	1	2	3	4	5	Competent
12	Ignorant	1	2	3	4	5	Knowledgeable
13	Irresponsible	1	2	3	4	5	Responsible
14	Unintelligent	1	2	3	4	5	Intelligent
15	Foolish	1	2	3	4	5	Sensible

Q2) Indicate the degree you agree with the following statements (1: strongly disagree, 5: strongly agree)

1 The robot acted intentionally[2]

1 2 3 4 5

2 The robot appeared to have the ability to make its own decisions

1 2 3 4 5

Q3) How many times have you had experience/interaction with a robot?

never 1 2 3 more

[2] This item was changed to "The robot gave intentional action commands' in Experiment 2.

Appendix B. Selected NimStim Stimuli used in Experiment 2

Appendix C. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2019.102819.

References

Barlas, Z., Hockley, W. E., & Obhi, S. S. (2017). The effects of freedom of choice in action selection on perceived mental effort and the sense of agency. Acta Psychologica, 180(September), 122–129. doi:10.1016/j.actpsy.2017.09.004.

Barlas, Z., Hockley, W. E., & Obhi, S. S. (2018). Effects of free choice and outcome valence on the sense of agency: Evidence from measures of intentional binding and feelings of control. Experimental Brain Research, 236(1), 129–139. doi:10.1007/s00221-017-5112-3.

Barlas, Z., & Kopp, S. (2018). Action choice and outcome congruency independently affect intentional binding and feeling of control judgments. Frontiers in Human Neuroscience, 12(April), 1–10. doi:10.3389/fnhum.2018.00137.

Barlas, Z., & Obhi, S. S. (2013). Freedom, choice, and the sense of agency. Frontiers in Human Neuroscience, 7(August), 514. doi:10.3389/fnhum.2013.00514.

Barlas, Z., & Obhi, S. S. (2014). Cultural background influences implicit but not explicit sense of agency for the production of musical tones. Consciousness and Cognition, 28(1), 94–103. doi:10.1016/j.concog.2014.06.013.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics, 1(1), 71–81. doi:10.1007/s12369-008-0001-3.

Berberian, B., Sarrazin, J.-C., Le Blaye, P., & Haggard, P. (2012). Automation technology and sense of control: A window on human agency. PloS One, 7(3), e34075. doi:10.1371/journal.pone.0034075.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. Trends in Cognitive Sciences, 6(6), 237–242.

Broadbent, E. (2017). Interactions with robots: the truths we reveal about ourselves. Annual Review of Psychology, 68(9), 1–926. doi:10.1146/annurev-psych-010416-043958.

Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. (2016). Coercion changes the sense of agency in the human brain. Current Biology, 26(5), 585–592. doi:10.1016/j.cub.2015.12.067.

Caspar, E. A., Cleeremans, A., & Haggard, P. (2018). Only giving orders? An experimental study of the sense of agency when giving or receiving commands. PLoS ONE, 13(9), e0204027. doi:10.1371/journal.pone.0204027.

Chambon, V., & Haggard, P. (2012). Sense of control depends on fluency of action selection, not motor performance. Cognition, 125(3), 441–451. doi:10.1016/j.cognition.2012.07.011.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale: Lawrence Erlbaum Associates Inc..

Cunnington, R., Windischberger, C., Deecke, L., & Moser, E. (2002). The preparation and execution of self-initiated and externally-triggered movement: A study of event-related fMRI. NeuroImage, 15(2), 373–385. doi:10.1006/nimg.2001.0976.

David, N., Newen, A., & Vogeley, K. (2008). The "sense of agency" and its underlying cognitive and neural mechanisms. Consciousness and Cognition, 17(2), 523–534. doi:10.1016/j.concog.2008.03.004.

Dewey, J. A., & Knoblich, G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing? PloS One, 9(10). doi:10.1371/journal.pone.0110118.

Duval, T. S., & Silvia, P. J. (2002). Self-awareness, prob- ability of improvement, and the self-serving bias. Journal of Personality and Social Psychology, 82(1), 49–61.

Ebert, J. P., & Wegner, D. M. (2010). Time warp: Authorship shapes the perceived timing of actions and events. Consciousness and Cognition, 19(1), 481–489. doi:10.1016/j.concog.2009.10.002.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160. doi:10.3758/BRM.41.4.1149.

Filevich, E., Vanneste, P., Brass, M., Fias, W., Haggard, P., & Kühn, S. (2013). Brain correlates of subjective freedom of choice. Consciousness and Cognition, 22(4), 1271–1284. doi:10.1016/j.concog.2013.08.011.

Frith, C. D. (2005). The self in action: Lessons from delusions of control. Consciousness and Cognition, 14(4), 752–770. doi:10.1016/j.concog.2005.04.002.

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. Philosophical Transactions of the Royal Society of London: Biological Sciences, 355(1404), 1771–1788. doi:10.1098/rstb.2000.0734.

Fukushima, H., Goto, Y., Maeda, T., Kato, M., & Umeda, S. (2013). Neural substrates for judgment of self-agency in ambiguous situations. PloS One, 8(8), e72267. doi:10.1371/journal.pone.0072267.

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. Trends in Cognitive Sciences, 4(1), 14–21.

Haggard, P. (2008). Human volition: Towards a neuroscience of will. Nature Reviews. Neuroscience, 9(12), 934–946. doi:10.1038/nrn2497.

Haggard, P. (2017). Sense of agency in the human brain. Nature Reviews Neuroscience. doi:10.1038/nrn.2017.14.

Haggard, P., & Chambon, V. (2012). Sense of agency. Current Biology: CB, 22(10), R390–R392. doi:10.1016/j.cub.2012.02.040.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. Nature Neuroscience, 5(4), 382–385. doi:10.1038/nn827.

Haggard, P., & Tsakiris, M. (2009). The experience of agency: Feelings, judgments, and responsibility. Current Directions in Psychological Science, 18(4), 242–246. doi:10.1111/j.1467-8721.2009.01644.x.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2), 65–70.

Hommel, B. (1996). S-R compatibility effects without response uncertainty. The Quarterly Journal of Experimental Psychology Section A, 49(3), 546–571. doi:10.1080/713755643.

Imaizumi, S., & Tanno, Y. (2019). Intentional binding coincides with explicit sense of agency. Consciousness and Cognition, 67(July 2018), 1–15. doi:10.1016/j.concog.2018.11.005.

Khalighinejad, N., Bahrami, B., Caspar, E. A., & Haggard, P. (2016). Social transmission of experience of agency: An experimental study. Frontiers in Psychology, 7(AUG), 1–11. doi:10.3389/fpsyg.2016.01315.

Khalighinejad, N., Di Costa, S., & Haggard, P. (2016). Endogenous action selection processes in dorsolateral prefrontal cortex contribute to sense of agency: A meta-analysis of tDCS studies of "intentional binding". Brain Stimulation, 9(3), 372–379. doi:10.1016/j.brs.2016.01.005.

Kühn, S., Brass, M., Haggard, P., Ku, S., Brass, M., Haggard, P., & Kühn, S. (2012). Feeling in control: Neural correlates of experience of agency. Cortex, 49(7), 1935–1942. doi:10.1016/j.cortex.2012.09.002.

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to intention. Science (New York, N.Y.), 303(5661), 1208–1210. doi:10.1126/science.1090973.

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2006). On measuring the perceived onsets of spontaneous actions. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 26(27), 7265–7271. doi:10.1523/JNEUROSCI.1138-06.2006.

Limerick, H., Coyle, D., & Moore, J. W. (2014). The experience of agency in human-computer interactions: A review. Frontiers in Human Neuroscience, 8(August), 643. doi:10.3389/fnhum.2014.00643.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? Psychological Bulletin, 82(2), 213–225. doi:10.1037/h0076486.

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. Consciousness and Cognition, 21(1), 59–68. doi:10.1016/j.concog.2011.08.010.

Moore, J. W., & Haggard, P. (2008). Awareness of action: Inference and prediction. Consciousness and Cognition, 17(1), 136–144. doi:10.1016/j.concog.2006.12.004.

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: A review. Consciousness and Cognition, 21(1), 546–561. doi:10.1016/j.concog.2011.12.002.

Moore, J. W., Ruge, D., Wenke, D., Rothwell, J., & Haggard, P. (2010). Disrupting the experience of control in the human brain: Pre-supplementary motor area contributes to the sense of agency. Proceedings. Biological Sciences/The Royal Society, 277(1693), 2503–2509. doi:10.1098/rspb.2010.0404.

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. Consciousness and Cognition, 18(4), 1056–1064. doi:10.1016/j.concog.2009.05.004.
Moreton, J., Callan, M. J., & Hughes, G. (2017). How much does emotional valence of action outcomes affect temporal binding? Consciousness and Cognition, 49, 25–34. doi:10.1016/j.concog.2016.12.008.
Moretto, G., Walsh, E., & Haggard, P. (2011). Experience of agency and sense of responsibility. Consciousness and Cognition, 20(4), 1847–1854. doi:10.1016/j.concog.2011.08.014.
Obhi, S. S., & Hall, P. (2011). Sense of agency and intentional binding in joint action. Experimental Brain Research, 211(3–4), 655–662. doi:10.1007/s00221-011-2675-2.
Obhi, S. S., & Hall, P. (2011). Sense of agency in joint action: Influence of human and computer co-actors. Experimental Brain Research, 211(3–4), 663–670. doi:10.1007/s00221-011-2675-2.
Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. Journal of Neuroscience Methods, 162(1–2), 8–13. doi:10.1016/j.jneumeth.2006.11.017.
Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. Frontiers in Neuroinformatics, 2. doi:10.3389/neuro.11.010.2008.
Poonian, S. K., & Cunnington, R. (2013). Intentional binding in self-made and observed actions. Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale, 229(3), 419–427. doi:10.1007/s00221-013-3505-5.
Sahaï, A., Desantis, A., Grynszpan, O., Pacherie, E., & Berberian, B. (2019). Action co-representation and the sense of agency during a joint Simon task: Comparing human and machine co-agents. Consciousness and Cognition, 67(March 2018), 44–55. doi:10.1016/j.concog.2018.11.008.
Sahaï, A., Pacherie, E., Grynszpan, O., & Berberian, B. (2017). Predictive mechanisms are not involved the same way during human-human vs. human-machine interactions: A review. Frontiers in Neurorobotics, 11(October). doi:10.3389/fnbot.2017.00052.
Satake, S., Hayashi, K., Nakatani, K., & Kanda, T. (2015). Field trial of an information-providing robot in a shopping mall. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015, 1832–1839. doi:10.1109/IROS.2015.7353616.
Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. Cognition, 94(3), 241–255. doi:10.1016/j.cognition.2004.04.003.
Schwartz, B. (2012). Choice, freedom, and autonomy. Meaning, Mortality, and Choice: The Social Psychology of Existential Concerns, 271–287.
Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? Cognition, 88(3), B11–B21. doi:10.1016/S0010-0277(03)00043-X.
Sebanz, N., & Lackner, U. (2007). Who's calling the shots? Intentional content and feelings of control. Consciousness and Cognition, 16(4), 859–876. doi:10.1016/j.concog.2006.08.002.
Sidarus, N., Vuorre, M., & Haggard, P. (2017). How action selection influences the sense of agency: An ERP study. NeuroImage, 150(February), 1–13. doi:10.1016/j.neuroimage.2017.02.015.
Simon, J. R. (1969). Reactions toward the source of stimulation. Journal of Experimental Psychology, 81(1), 174–176. doi:10.1037/h0027448.
Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. Journal of Applied Psychology, 51(3), 300–304. doi:10.1037/h0020586.
Sperduti, M., Delaveau, P., Fossati, P., & Nadel, J. (2011). Different brain structures related to self- and external-agency attribution: A brief review and meta-analysis. Brain Structure & Function, 216(2), 151–157. doi:10.1007/s00429-010-0298-1.
Stenzel, A., Chinellato, E., Bou, M. A. T., del Pobil, Á., Lappe, M., & Liepelt, R. (2012). When humanoid robots become human-like interaction partners: Corepresentation of robotic actions. Journal of Experimental Psychology: Human Perception and Performance, 38(5), 1073–1077. doi:10.1037/a0029493.
Stenzel, A., Dolk, T., Colzato, L. S., Sellaro, R., Hommel, B., & Liepelt, R. (2014). The joint Simon effect depends on perceived agency, but not intentionality, of the alternative action. Frontiers in Human Neuroscience, 8(August), 1–10. doi:10.3389/fnhum.2014.00595.
Strother, L., House, K. A., & Obhi, S. S. (2010). Subjective agency and awareness of shared actions. Consciousness and Cognition, 19(1), 12–20. doi:10.1016/j.concog.2009.12.007.
Takahata, K., Takahashi, H., Maeda, T., Umeda, S., Suhara, T., Mimura, M., & Kato, M. (2012). It's not my fault: Postdictive modulation of intentional binding by monetary gains and losses. PloS One, 7(12), e53421. doi:10.1371/journal.pone.0053421.
Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: separating fact from fiction. Psychological Bulletin, 116(1), 21–27. discussion 28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8078971.
Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: Judgements from untrained research participants. Psychiatry Research, 168(3), 242–249. doi:10.1016/j.psychres.2008.05.006.
van der Woerdt, S., & Haselager, P. (2017). When robots appear to have a mind: The human perception of machine agency and responsibility. New Ideas in Psychology, (November), 1. doi:10.1016/j.newideapsych.2017.11.001.
Waszak, F., Wascher, E., Keller, P., Koch, I., Aschersleben, G., Rosenbaum, D. A., & Prinz, W. (2005). Intention-based and stimulus-based mechanisms in action selection. Experimental Brain Research, 162(3), 346–356. doi:10.1007/s00221-004-2183-8.
Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. American Psychologist, 54(7), 480–492.
Wen, W., Yamashita, A., & Asama, H. (2015). The influence of action-outcome delay and arousal on sense of agency and the intentional binding effect. Consciousness and Cognition, 36, 87–95. doi:10.1016/j.concog.2015.06.004.
Wenke, D., Fleming, S. M., & Haggard, P. (2010). Subliminal priming of actions influences sense of control over effects of action. Cognition, 115(1), 26–38. doi:10.1016/j.cognition.2009.10.016.
Wohlschläger, A., Engbert, K., & Haggard, P. (2003). Intentionality as a constituting condition for the own self and other selves. Consciousness and Cognition, 12(4), 708–716. doi:10.1016/S1053-8100(03)00083-7.
Wolpert, D. M. (1997). Computational approaches to motor control. Trends in Cognitive Science, 1(6), 209–216. doi:10.1016/S1364-6613(97)01070-X.
Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. Science, 269(5232), 1880–1882. doi:10.1126/science.7569931.
Yoshie, M., & Haggard, P. (2013). Negative emotional outcomes attenuate sense of agency over voluntary actions. Current Biology, 1–5. doi:10.1016/j.cub.2013.08.034.
Yoshie, M., & Haggard, P. (2017). Effects of emotional valence on sense of agency require a predictive model. Scientific Reports, 7(1), 8733. doi:10.1038/s41598-017-08803-3.